ENHANCEMENT OF DIFFERENTIALLY CODED SPEECH SIGNALS BY MEANS OF QUANTIZATION ERROR CORRECTION

P.J.J.A. Wolters

Twente University of Technology, P.O. Box 217,
7500 AE ENSCHEDE. The Netherlands.

ABSTRACT

A method to increase the Signal-to-Noise Ratio (SNR) of speech codecs is presented. It has been applied to Adaptive Delta Modulation (ADM) according to Jayant's algorithm.
Observations indicate that bit streams resulting from these coders and quantization errors are correlated to some extend. Based on this experience a Quantization Error Correction method (QEC) has been developed that improves the SNR of the output signal. With this method the decoder output is updated with a correction signal derived from the received bit sequence.
An expression for the optimal correction level is derived. Application of the method by means of computer simulation and hardware implementation gave a 5 to 6 dB increase in SNR.
There is no need to transmit additional information and a normal decoder can be used together with additional circuitry.

INTRODUCTION

In spite of the well-accepted reasons for digitizing speech signals, the large bandwidth required is a major drawback. Therefore at our institute a research program has been started on the improvement of differential speech encoding techniques [1,2]. This paper deals with some intermediate results and has not the status of a final report.
Within the scope of the project different coding algorithms have been implemented in hardware and/or software. Using the commonly known Adaptive Delta Modulation principle (ADM) according to Jayant [3,4] certain regularities concerning the relation between quantization errors and output bits have been perceived. Referring to fig.1 the ADM algorithm is given by:

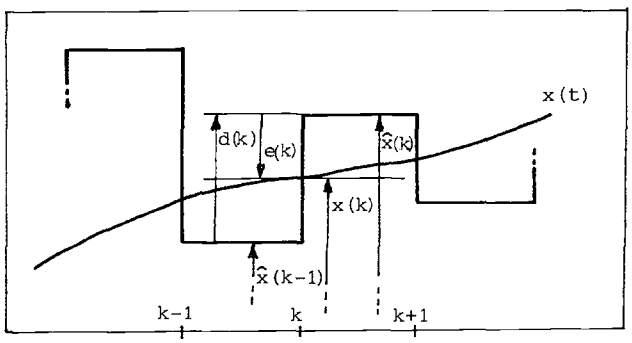

Fig. 1 Definition of ADM symbols

$$\text{output bit: } c(k) = \frac{1}{2}\{\,1 + \operatorname{sgn}[x(k) - \hat{x}(k-1)]\,\}$$

$$\underline{\text{if}} \quad c(k) = c(k-1) \ \underline{\text{then}}\ d(k) = P.d(k-1)$$
$$\underline{\text{else}}\ d(k) = -Q.d(k-1)$$
$$\text{with } P = 1/Q = 1.5$$

$$\hat{x}(k) = \hat{x}(k-1) + d(k)$$

The quantization error on sample moment k is given as $e(k)=x(k)-\hat{x}(k)$.
We achieved an increase in Signal-to-Noise Ratio (SNR) through minimization of the Mean-Square Quantization Error (MSQE). We are aware that a coding method being optimal in this sense is not always experienced optimally in subjective tests.

EXPLANATION OF ENHANCEMENT PRINCIPLE.

Observations of ADM coded speech signals showed that the resulting bit streams are not totally decorrelated. This indicates that information is left in the digital output signal of ADM coders. The following example illustrates the kind of effects that can be observed. Assume the occurrence of a slope-overload situation. In this situation long sequences of "ones" or "zeros" appear. Moreover the sign of the quantization error during such a sequence is predictable because a series of "ones" is accompanied by a series of positive quantization errors. Comparable effects occurring during situations other than slope-overload induced us to investigate the statistical properties of quantization errors and their specific relation to the output bit sequence of the coder. If the existence of this relation can be proved it is possible to utilize this knowledge at the decoder in order to improve the MSQE. It was the aim to use a normal codec together with some additional correction algorithm. So there is no anticipated change in the codec and no additional information has to be transmitted.
In order to investigate statistical properties of e(k) in relation to the occurrence of certain bit sequences, bit combination i on sample moment k is defined as the (2N+1)-bit binary word:

$$c(k-N), \ldots, c(k-1), c(k), c(k+1), \ldots, c(k+N)$$

We will further refer to i as the decimal representation of the respective binary word, e.g.

CH1841-6/83/0000-1296 $1.00 © 1983 IEEE

i=6 means bit sequence 110.
The power of the quantization error is written as:

$$E\{e^2(k)\} = m^2\{e(k)\} + var\{e(k)\}$$

in which m{e(k)} and var{e(k)} are the mean and the variance of e(k). Normally m{e(k)} is zero, however the mean value of e(k), determined under the condition of a certain bit sequence i: $m_i\{e(k)\}$, will generally not be zero. The power of the quantization error for each bit combination i is:

$$E_i\{e_i^2(k)\} = m_i^2\{e_i(k)\} + var_i\{e_i(k)\}$$

If $m_i\{e(k)\}$ is not zero, it can be forced to zero by adding the value $m_i\{e(k)\}$ to each decoded output level at bit combination i. The corrected output level can be written as:

$$\hat{x}c_i(k) = \hat{x}_i(k) + m_i\{e_i(k)\}$$

and so:

$$E_i\{ec_i^2(k)\} = E\{(x_i(k) - \hat{x}c_i(k))^2\} = var_i\{e_i(k)\}$$

For each bit combination the quotient of the SNR after and before the correction can be written as:

$$G_i = 1 + \frac{m_i^2\{e_i(k)\}}{var_i\{e_i(k)\}}$$

This quotient, called gain in SNR for bit combination i, is greater than one for all values of $m_i\{e(k)\}$ and $var_i\{e(k)\}$.
Due to the large dynamic range of speech signals the quantization error has a comparably large dynamic range. So unfortunately the variance in the error is also large and this usually makes G_i relatively small. The adaptation algorithm makes the step magnitude more or less proportional to the amplitude of the input signal. Therefore a normalized quantization error was defined as:

$$ne(k) = \frac{e(k)}{d(k)}$$

and the correction method was changed accordingly into:

$$\hat{x}c_i(k) = \hat{x}_i(k) + m_i\{ne_i(k)\}.d_i(k)$$

In this approach it is not possible to give a simple expression for G_i from which it can be concluded that it will be larger than one for various speech signals. In order to get some understanding of this error correction algorithm it was tried to estimate the probability density function (pdf) of $ne_i(k)$. In the case of N=1 (a three bit correction algorithm) a simple model for the ADM process has been developed, based on several assumptions. One of these is that the speech signal is oversampled which is valid for normal ADM coders. Another assumption is that the normalized error is limited in amplitude as follows:

$$-1 \leq ne(k) \leq 1$$

This is very likely if the speech signal is reasonably well tracked. In table I the values for $m_i\{ne(k)\}$ based on the model are given.

i	$m_i\{ne(k)\}$	i	$m_i\{ne(k)\}$
0	0,20	4	0,13
1	-0,70	5	-0,52
2	-0,52	6	-0,70
3	0,13	7	0,20

Table I: $m_i\{ne(k)\}$ for N= 1

Some of these values differ significantly from zero which is encouraging to pursue this approach.

OPTIMAL ENHANCEMENT METHOD.

If it is assumed that the output signal of an ADM decoder is corrected with a signal $a_i.d_i(k)$ in which a_i is a constant for each bit combination so that:

$$\hat{x}c_i(k) = \hat{x}_i(k) + a_i.d_i(k)$$

and a_i is heuristically chosen to be equal to $m_i\{ne(k)\}$ it cannot be expected that the SNR is optimal in the sense of minimal MSQE. In the expression above a_i is not chosen to be dependent on the actual d(k) for reasons of simplicity in implementation.
The value for a_i that produces an optimal SNR is derived as follows. The power of the quantization noise for each bit combination is:

$$N_i = E_i\{e_i^2(k)\} = E_i\{(x_i(k) - \hat{x}_i(k))^2\}$$

and in the case of a corrected output signal:

$$Nc_i = E_i\{ec_i^2(k)\} = E_i\{(e_i(k) - a_i.d_i(k))^2\}$$

This noise power is minimal if its derivative is zero:

$$2E_i\{e_i(k).d_i(k)\} - 2a_i.E_i\{d_i^2(k)\} = 0$$

or:

$$(a_i)_{opt} = \frac{E_i\{e_i(k).d_i(k)\}}{E_i\{d_i^2(k)\}}$$

Substitution of the optimal a_i in the expression for Nc gives:

$$(Nc_i)_{min} = E_i\{e_i^2(k)\} - \frac{E_i^2\{e_i(k).d_i(k)\}}{E_i\{d_i^2(k)\}}$$

and for the gain in SNR can now be written:

$$G_i = \frac{E_i\{e_i^2(k)\}.E_i\{d_i^2(k)\}}{E_i\{e_i^2(k)\}.E_i\{d_i^2(k)\} - E_i^2\{e_i(k).d_i(k)\}}$$

With the described Quantization Error Correction method (QEC) this gain is always greater than 1 [5] and we may undeniably speek of an enhancement method. To verify this theoretical result a computer simulation has been performed and a hardware prototype built.

COMPUTER SIMULATIONS AND MEASUREMENTS.

Rather long utterances of male and female voices have been used for the computer simulations in order to obtain statistically stable results. The typical length used was 8-16 secs. The input signal had an amplitude of 10 Vp-p and was sampled with various frequencies and A-D converted with 12 bit accuracy. An input low-pass filter was applied with a cut-off frequency of 3400 Hz. For a large number of different speech segments the pdf of $ne_i(k)$ was determined. These pdf's are very much like the ones that were based on the ADM model for N=1 mentioned before. The experiments showed further that for several values of i $m_i\{ne_i(k)\}$ differs significantly from zero. These results were also more or less independent of the speech segments used. The pdf of ne(k) turns out to be approximately uniform for $-1<ne(k)<+1$ and about 96% of all normalized errors are in this region.
The correction algorithm based on $m_i\{ne_i(k)\}$ has been implemented. An improvement in SNR of 1 to 5 dB has been observed for sampling frequencies (f_s) of 16 and 32 kHz. These preliminary simulations are of limited significance and hence will only be discussed briefly.
More important results have been achieved through the use of the optimal QEC method. The correction factors a_i have been determined for various speech segments. An example of these factors for a male voice in the case of N=1 and f_s=16 and 32 kHz is given in table II.

i	a_i,16kHz	a_i,32kHz	i	a_i,16kHz	a_i,32kHz
0	0,30	0,10	4	0,38	0 40
1	-0,61	-0,72	5	-0,48	-0,44
2	-0,49	-0,44	6	-0,60	-0,72
3	0,41	0,39	7	0,30	0,10

Table II: a_i for N=1, male voice, f_s= 16 and 32 kHz

It can be concluded that these correction factors have considerable values. Though not presented here, we have seen that this is also the case for N=2,3 and 4. Comparison with table I shows that the signs of a_i and $m_i\{ne_i(k)\}$ agree while corresponding a_i factors have about the same magnitude. The values for both sampling frequencies differ sligthly.
The optimal QEC algorithm has also been applied to various speech segments. Fig.2 gives the gain in SNR, expressed in dB's for two sampling frequencies and four values of N in the case of a male voice. This gain is about the same for different male and female voices.
The gain in SNR has been calculated by subtracting the (log) power of the corrected error signal from the uncorrected one, using

$$P(e) = 10 * \log\{\frac{1}{K}\sum_{k=1}^{K} e^2(k)\}$$

It should be stated here that the error signal has not been low-pass filtered so that results from fig.2 might be somewhat flattered.

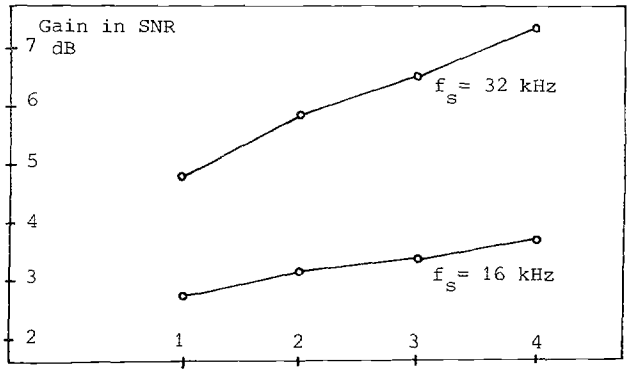

Fig. 2 Gain in SNR, male voice, f_s= 16 and 32 kHz

HARDWARE IMPLEMENTATION OF QEC ALGORITHM.

It is preferable that speech encoding algorithms are evaluated by means of subjective listening tests and for this reason a hardware prototype has been constructed. The block diagram is given in fig.3.

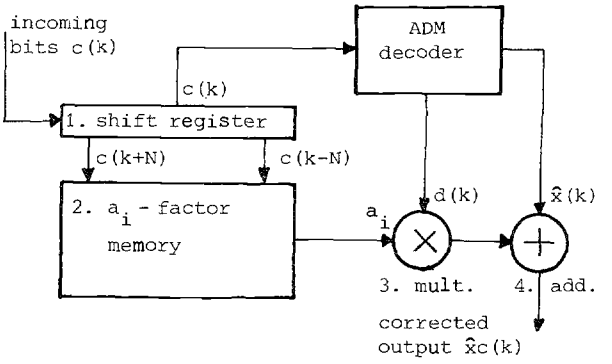

Fig. 3 Block diagram of ADM decoder with QEC.

In addition to a normal (all-digital) decoder we find:
1. a (2N+1)-bit shift register containing bit combination i. The bit in the middle is the one used in the decoder.
2. an a_i-factor memory consisting of (2N+1) a_i-factors. This memory has been made with EPROM's which are filled with a_i-factors of three bits before and three bits after the decimal point. They have been calculated by means of a computer from a female speech utterance that was sampled with 16 kHz. This implies that for both this calculation and for the measurements to be described, the exact same speech segment is not used, and mostly not even the same voice.
3. a digital multiplier to carry out the multiplication of a_i and $d_i(k)$.
4. a digital adder in which the correction value is added to the output reconstruction level.
The system has been equipped with two D-A converters so that the uncorrected as well as the corrected signal can be monitored at the same time. Three SNR-measurement methods have been applied for evaluation purposes. For the first two methods a hardware measurement system has been used [6] to measure

$$var\{e(k)\} = E\{e^2(k)\}$$

With this system it is possible to subtract the input signal of a codec from its output signal. This is achieved through delaying the input signal in a high resolution digital delay circuit by an amount at which the power of the error signal is minimal. Both var{e(k)} and var{x(k)} are calculated and the SNR is presented. With this method SNR figures are achieved that differ significantly from those obtained through computer measurements. The reason for this is that in a computer normally no compensation for signal delay in the codec over a fraction of the sampling period is made. The three methods applied are:

1. SNR measurement with sine-wave input signals. In the frequency range of 300 to 3000 Hz typical values for the gain in SNR are 1.3 dB at f_s=16 kHz and 1.5 dB at f_s=32 kHz. Since the a_i-factors used are not optimal for sine-waves no significant gain could be expected.
2. SNR measurement with speech input signals. Some results of these measurements are given in table III.

	f_s= 16 kHz	f_s= 32 kHz
male voice	2,7 dB	0,8 dB
female voice	3,8 dB	0,5 dB

Table III: Gain in SNR, with a_i determined from female voice; f_s= 16 kHz

The fact that the a_i-factors used have been determined from female speech with f_s=16 kHz most likely explains why the gain for another female voice is higher than for a male voice and that better results occur at sampling frequencies of 16 kHz.

3. SNR measurement with noise input signals. With this method band-filtered white noise in the frequency range of 450 to 550 Hz is used as stimulus for the codec. The output noise in the frequency range from 850 to 3400 Hz is measured with a noise level meter. Fig.4 gives the result of this measurement: with various input levels a significant improvement in SNR is achieved.

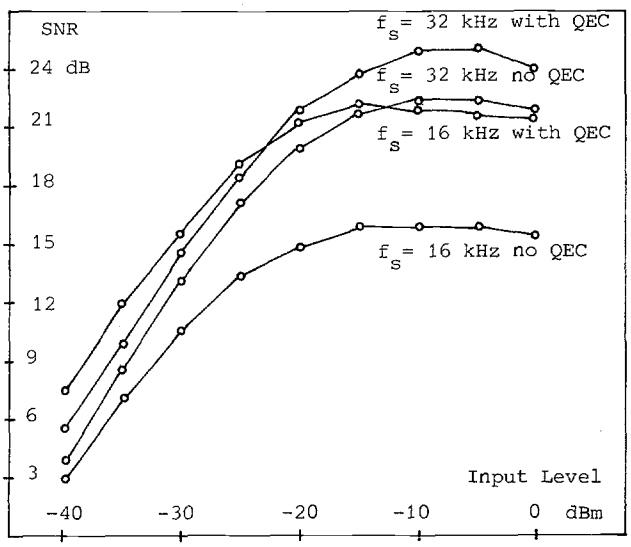

Fig. 4 SNR as a function of input level, N= 4, noise input signal.

DISCUSSION

In the computer simulations the QEC algorithm has been applied to speech segments using a_i-factors originating from exactly the same speech segments. In the hardware prototype different measurements have been executed with only one set of a_i-factors belonging to a female speech utterance. This explaines why results from simulations and hardware measurements differ.

It can further be seen that contrary to the outcome of simulations the gain in SNR for the hardware prototype becomes lower at higher sampling frequencies. This is probably caused by the absence of a low-pass filter in the simulations. Apparently the QEC algorithm has a considerable effect on frequencies above 3400 Hz also.

The first results with a QEC algorithm have been presented but further research concerning the influence of the low-pass filter and of other coding parameters such as P and Q will be carried out. The influence of deviations from the optimal a_i-factors on the gain in SNR will also be investigated.

It seems very likely that a comparable form of QEC can succesfully be applied to other differential encoding techniques. We have recently performed some experiments with Delta Modulated images and observed a gain in SNR of up to 10 dB.

ACKNOWLEDGEMENT

The author wishes to thank the colleagues and students who helped him prepare this paper.

REFERENCES

1. N.S.Jayant,"Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers",proc.IEEE,vol 62,No 5,May '74,pp 611-632.
2. J.L.Flanagan et.al,"Speech Coding",IEEE trans.on comm.,vol COM-27,No 4,April '79,pp 710-736.
3. N.S.Jayant,"Adaptive Delta Modulation with a One-bit Memory",Bell.Syst.Tech.J,vol 49,No 3, March '70,pp 321-342.
4. R.Steele,"Delta Modulation Systems" ,London, Pentech Press,1975.
5. A.Papoulis,"Probability, Random Variables and Stochastic Processes",Tokyo ,McGraw-Hill Koga-kusha Ltd ,1965.
6. P.J.J.A.Wolters,"A Microprocessor-Controlled Signal-to-Noise Ratio Measurement System for Speech Codecs",IEEE trans.on Instr.and Meas,vol IM-31,No 1,March '82,pp 12-17.